



**Manchester
Metropolitan
University**

Talukder, Reja Alam, Ahsan, Mominul and Andersson, Karl (2015) Visualization and Cell Data Analysis Tool based on XML Log Files. In: 2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops), 26 October 2015 - 29 October 2015, Clearwater Beach, FL, USA.

Downloaded from: <https://e-space.mmu.ac.uk/624871/>

Publisher: IEEE

DOI: <https://doi.org/10.1109/LCNW.2015.7365941>

Please cite the published version

<https://e-space.mmu.ac.uk>

Visualization and Cell Data Analysis Tool based on XML Log Files

Reja Alam Talukder and Karl Andersson

Pervasive and Mobile Computing Laboratory
Luleå University of Technology
S-931 87 Skellefteå, Sweden

Mominul Ahsan

Faculty of Engineering and Computing
Dublin City University
Glasnevin, Dublin 9, Ireland

Abstract—Mobility management is an essential feature of cellular networks. High accuracy of mobile user positioning is needed to handle mobility efficiently enough and bad cell data can harm this feature significantly. Inaccuracy of cell shapes, lack of cell data measurements, and inaccurate coordination in a geographical area are major shortcomings when it comes to positioning of mobile users in cellular networks. This paper describes a tool that visualizes and analyzes cell data based on XML log files. The tool evaluates a mathematical expression to identify bad cells from the log file and successfully fixes most of the bad cells identified. The tool repairs bad cell shapes in order to achieve better positioning of mobile users.

Keywords— XML log files; Visualization tool; Cell data; Mobile user positioning; Cellular networks

I. INTRODUCTION

Personal communication systems are developing with the development of technologies. Achieving better mobile user's position is still controversial. Mobile operators are still searching to provide better services through different technological advancement. Recently, GSM and CDMA mobile operators are trying to develop different modern technology to provide better solutions for the mobile user location purposes. Cellular data is distributed in logical geographical areas while cell sites generate a frequency radiator, which is covered by logical geo-coordinate system. Each geo-coordinate system has a certain measurement within a certain area. The cell shapes of cellular data have different sizes such as square, circular and hexagonal etc. According to the cell data analysis real cell data are found as inappropriate size. XML Log file contains cellular data to store the mobile positioning data. Latitude and longitude data have been considered from the log file to analysis accuracy of cell data shape. Due to the inaccuracy of cell data shape mobile operators are in difficulties to find user position perfectly. According to the cell data definition, a good cell is logically defined in a geo-coordinate area and mobile phones are located within perfect range of latitude and longitude of that cell data. The latitude and longitude of any two points are calculated approximately seventy km distance for a good cell data. On the contrary, longer distance within latitude and longitude of two

cell points are defined as bad cell data. Therefore, user position cannot be located properly. Mobile operators are observing irregular cell data shape as well as difficulties for finding mobile user location in GSM and CDMA infrastructure. The proposed idea can identify those bad cell data and fix into good cell data due to better accuracy of a user position.

The paper describes study background in section 2. The proposed system architecture and solution has been described in section 3. Experimental works are presented in the section 4 and section 5 describes the experimental results. Finally conclusion remark is placed in section 6.

II. BACKGROUND

Data mining is a process to identify patterns and characteristics of a particular type of data set. A data mining technique with XML has been established to understand the patterns of positioning data. Classification, estimation, prediction clustering etc. are performed in data mining technique to analyze positioning data. There are different types of technologies are used in the data mining technique. Statistical analysis is one of the important branches in data mining technique. Mathematical analysis is required to conduct data mining. In the proposed system, pattern recognition has been used to understand the shape of mobile cell data [1].

Mengjie et al. [2] used a data mining technique to complete building indexes using MapReduce. Heap algorithms have been used to resort XML parsed files. SAX and DOM XML parsing tools are used to analyze complicated data structure in their system. However, the system does not analysis any cell data shape. Eric et al. [3] proposed an approach to estimate position of base station out of crowd-sourced data using open mobile network. Semantic principles of linked data are used to publish mobile network. The grid-based approach is conducted as a result to estimate the base station position by comparing with other locations. The system does not indicate improvement of the cell data shape for user positioning purposes. Ricardo et al [4] proposed a solution to query and explore XML documents. The tool can able to manipulate XML document by offering innovative features of user communication using fisheye trees view and visualizing results. The system provides suitable interaction among users which makes different from other applications. However, the system is unable to provide information on cell data shape as well as

mobile user location identification. Jaymin et al. [5] proposed a prototype system to visualize network coverage map, identification of cell boundaries, cell level measurements etc. using mobile device parameters. Battery optimization algorithm is used to store and communicate with their system. All the measurement and information are updated in the server to process, store, and visualize data to identify patterns and trends of data. The system can identify and cover some areas affected by buildings and other obstacles. However, the system needs further investigation to improve cell data shape for user positioning purposes. H. Xu et al. [6] proposed a similarity-based Bayesian learning approach to diagnosis fault. It can estimate similarity degrees of structural elements from different log files, construct Bayesian network and to compute probabilities using similarity-based learning algorithm. Finally different structures of data sets are classified into fault categories. The paper describes fault diagnosis by analyzing semi-structured log data rather than cell data shape analysis and repair. Huixin et al. [7] proposed a location base service SUPL that supports GPS and cell id to determine the location base service in China. SUPL is designed by OMA including system architecture and network entities. GPS is used for positioning purposes and cell id is used to cover some geographical places. The service processes non roaming proxy mode although it is not described about visualization tool. Huang et al [8] has proposed a method of XML technology to parse log file by analyzing format of Vista event log file. Binary XML technology has been used to organize the data. The parsing output shows the event log file from XML format into textual XML format. The author has described the log file conversion based on windows vista. However, the system does not describe log file for mobile user positioning purposes. Xianjun et al. [9] introduced web log mining technology based on XML to recognize the user and conversation. The method is designed by star logic modelling and redundant dimension data to improve information index function. The system defines web log information in dimension to understand further mining for logging web data from user. The system describes web based mining information rather than finding specific user position. Wa et al [10] proposed a priority-based resource sharing scheme for integrating voice/data in cellular networks. The system serves high priority voice traffic, complex interaction of voice and data traffic, determine optimal CAC parameters, load adaption and bandwidth allocation adjustment. A mathematical model has been introduced to measure complex interaction between the voice and data traffic shared their total resources. However the system focuses more on internet data services. The author has described the characteristics of the XML standard for representing web-based data. The paper also shows different types of data integration techniques. Elisa et al [11] described the characteristics of XML and data integration procedure rather than providing any XML data mining tool to extract positioning data. Chiris et al [12] proposed a cell shape based system architecture to find user position by WAP. A simulation has been proposed based on mobile positioning that can locate the mobile station within a certain distance. A position algorithm has been developed to receive signal from

neighboring cells. The system can only able to provide better user position calculating a short range rather than covering widely. The performance comparison between the proposal and the other relevant existing schemes is described as follows.

Yang et al [13] analyzed causes for large localization errors and proposed a bounding technique to select area of wardriving trace and studied its performance. Finally, they identified a cell combining optimization to reduce localization errors by half. Our proposed system can improve cell data accuracy for position purposes. Ulm et al [14] used OpenCellID database and tried to estimate the position of mobile devices along with the positioning accuracy. However, our proposed system improves cell data shape for better user location purposes. Neidhardt et al. [15] described a grid-based approach for estimating better accuracy of base station positions compared with their real locations. An OpenMobileNetwork has been used to visualize the base station positions on the OpenMobileNetwork Map. However, our proposed approach converts most of the bad cell into good cell data for getting higher accuracy of user location. Mikami et al. [16] proposed cell-identification performance improvement system with transmission power coordination of the synchronization signal (SS) symbols for HetNet with Cell Range Expansion (CRE) to evaluate the cell identification characteristics of a cell search operation for identifying pico-cell. On the contrary our proposed system can successfully search, visualize and identify bad cell data on Google map.

After a close critical analysis of literature background we concluded that the existing systems do not improve the cell data shape for better user positioning purposes. Therefore, there is still a need for further investigation on cell data and to improving cell data shape for providing better user position.

III. PROPOSED SOLUTION

A. XML data conversion

XML log file data is generated from BTS. The mobile operator receives the log file data from BTS via file transfer protocol (FTP). The XML log file is used for all kind of cell data information. A database has been created to store the data. Data analysis and reporting are conducted in data warehouse. Furthermore, raw cell data has been analyzed from relevant cell data and finally those are shown on Google map. Figure 1 shows all the steps involved in the XML data conversion.

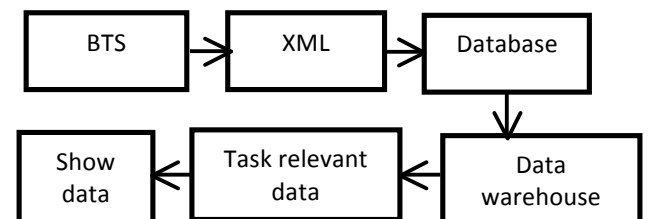


Fig. 1. XML data conversion.

B. Visualization tool algorithm

The mathematical idea behind the algorithm for managing cell data is based on the haversine formula ($\text{havrsin}(\theta) = \sin^2(\theta/2)$) to calculate longitude and latitude of cell data points. To measure the haversine of the central angle between two cell data points are calculated using $\text{Haversin}(d/r) = \text{havrsin}()$

$$\frac{d}{r} = \text{havrsin}(\phi_1 - \phi_2) + \cos(\phi_2) \text{havrsin}(\lambda_2 - \lambda_1), d \text{ is the distance between the two points, } \phi_1, \phi_2: \text{Latitude of two points, } \lambda_1, \lambda_2: \text{Longitude of two points, } d/r \text{ is central angle in radians. Furthermore, the distance between two data points have been considered as 70 km to draw the polygon which is defined as good cell. Furthermore, this bad cell data have been repaired into good cell data using the techniques of algorithm in the visualization tool, which has been explained in section 3 (see figure 2).$$

A database has been created to insert all kind of data to identify bad cell and good cell. Afterwards, XML is incorporated with the proposed algorithm and developed a visualization tool to extract data from the real data set collected from the telecom company. At first the visualization tool is run on Google map. The user presses click point on Google maps in any specific area. Some points from the click point on specific area are displayed. Those points value are matched with real cell data from the database. Afterwards, users get some point value from the click point in those specific area and cell data are visualized on Google maps. Those data indicate good and bad cell shapes. Refresh button is used to initialize the process if users are unable to find any data from click point in any specific area on Google maps. Finally Good cell data and bad cell data are shown on Google maps. Afterwards, bad cell data are filtered and considered for repairing by the proposed algorithm. Most of them are repaired into good cell data. Figure 2 shows the visualization algorithm.

Figure 2 shows the visualization algorithm. The process starts with 'Load Google map' and 'Click point'. A decision 'Marching point from DB' follows. If 'No', it leads to 'Refresh' and back to 'Click point'. If 'Yes', it leads to 'Show data'. Another decision 'Show good and bad data' follows. If 'Bad', it leads to 'Repairable data'. If 'Good', it leads to 'Good data'.

Fig. 2. Visualization algorithm.

C. Tool overview

The main goal of the proposed solution is to examine and display visualization data using programming and mathematical function. Therefore, the tool has been developed to collect and visualize cell data. Users can select a specific

geographical area for getting cell data. The positioning data has been compared with the data in the database. The data has been analyzed focusing the accuracy of user position. Furthermore, data has been identified as good and bad cell data according to Figure 3. Once the bad cell data are identified then the system can process to fix the bad cell into the good cell by changing their shape as hexagonal. It is seen that most of the cell shapes are displayed in an accurate form of data on Good map. Figure 3 describes the tool overview.

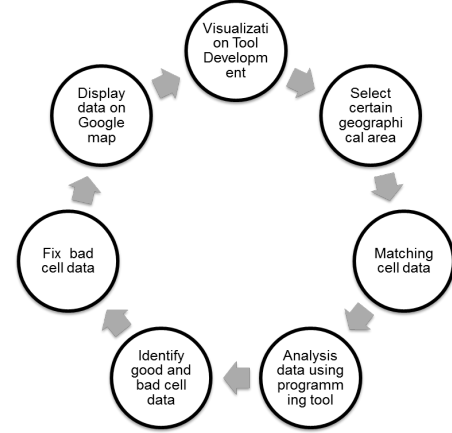


Fig. 3. Tool overview.

D. Software architecture of the visualization tool

XML log file data are collected from mobile operators. Cell data has been extracted from the log file. The database can store the formatted data. Internet connection has been established to communicate data between client and the database. The client can visualize the data from the client computer according to the algorithm. The data is shown on Google maps with their different cell shape and size. The user can recognize good and bad cell data based on regular and irregular cell shape of those positioning data. The tool can able to fix the irregular cell data shape in to regular cell data shape for better positioning purposes according to Figure 4.

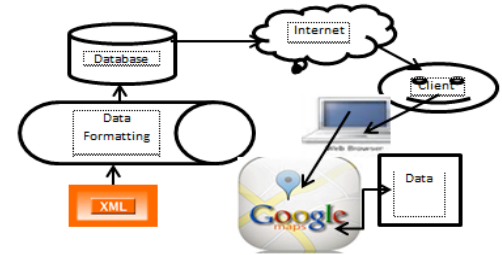


Fig. 4. Software architecture of visualization tool.

IV. EXPERIMENTAL ENVIRONMENT

A. Database

All data is stored simultaneously with all information from the log file according to Figure 5. The database contains few tables as cell table, celldata table, fix table, goodcells table to store different categories of data according to Figure 5. The

Cell table has stored for all bad cell data. The Celldata table contains all formatting positioning data. Consequently, the Fix table is used to store the fixed data after repairing.

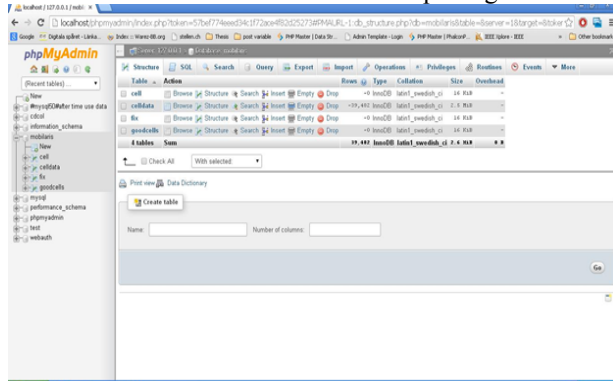


Fig. 5. Database of cell data.

B. Good cell data shape

Once the visualization tool is run, the good cell data shape are displayed on Google map as Figure 6. The system is able to find many good cell shapes. The view of a good cell is displayed in the figure. The cell size looks as hexagonal cell data shape. Therefore, according to the good cell definition it is claimed that it is good cell data.

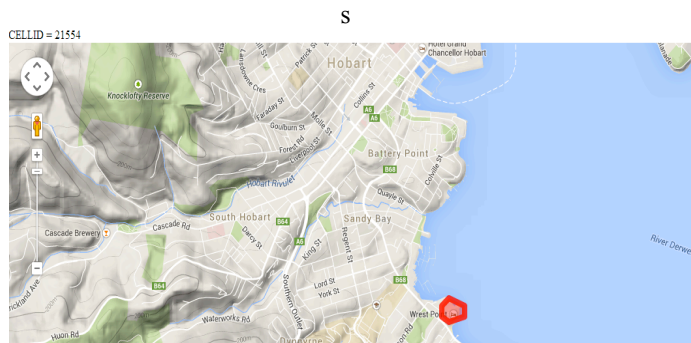


Fig. 6. Good cell data shape.

C. Bad cells

Figure 7 indicates the characteristics of a bad cell data. Once the tool is running the following cell shape is created after considering few points of cell on Google map. The cell shape is not regular shape as hexagonal. Therefore, it is considered as bad cell data shape.

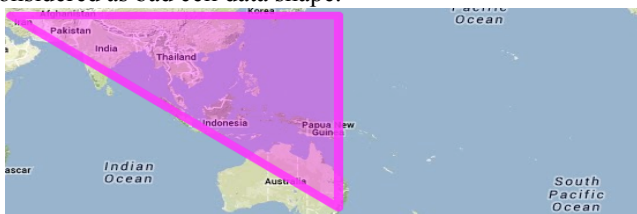


Fig. 7. Bad cell data shape.

V. RESULTS

A. Bad and good cell data

The bad cell data and good cell data with their shape after running visualization tool on Google map is shown in Figure 8.

User can find all cell data for a particular geographical area on Google maps. Process button in the tool can process the selected area on the map. Afterwards, the data and cell shape on Google maps are displayed using the click button on the tool. Furthermore, repair button can fix the bad cell data once the repairable data are available. There are also a few color buttons on the tool to see the cell shape in different color.

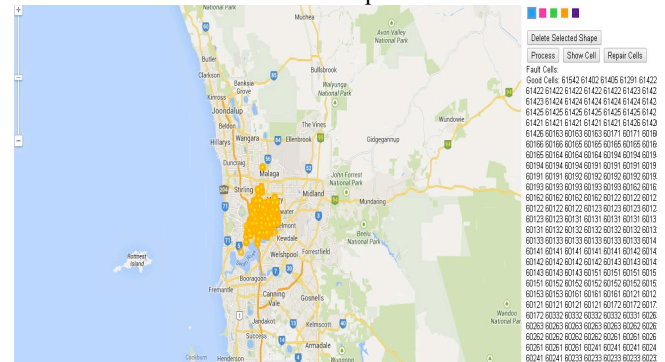


Fig. 8. Good cell data with shape.

B. Zoom view of bad and good cell data

Figure 9 shows the cell shape with data after zooming in the tool. Most of the cell shapes look hexagonal sizes which indicate as good cell. The color of the shapes displayed here as yellow in the figure. User can change the color from the different color options on the right top corner in the tool. The software tool also proposed all information of cell data in the user interface.

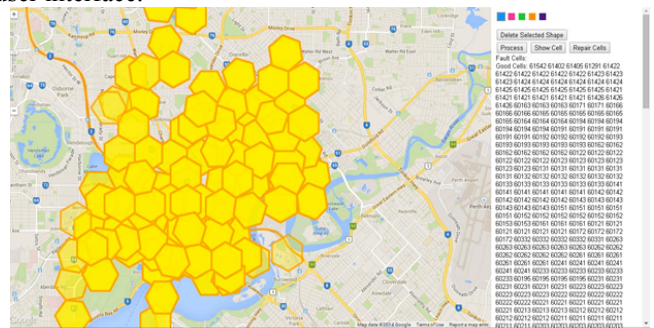


Fig. 9. Good cell data shape (zoomed).

C. Repair cell data

The repairable bad cell data is considered for repair on Google maps using the tool shown in Figure 10. The algorithm is used to identify good cell shape from the bad cell data within a range of seventy km distance between two points based on latitude and longitude of cell data. Few points of a cell data are taken as an experimental test to convert as good cell shape. As example of a cell point 10.2, 10.3, 10.4 and 200.4 km are assumed to conduct the experiment using the tool. First three

values are nearly similar and the last value is relatively higher than others. The tool can able to replace a value as first three values instead of 200.4 to make all the points similar. Therefore, a good cell shape is converted from the bad cell data. The approach has been used to convert majority of bad cell data into good cell data considering different cell data. However, a few bad cell data containing a longer distance data points are unmanageable to fix with this system.

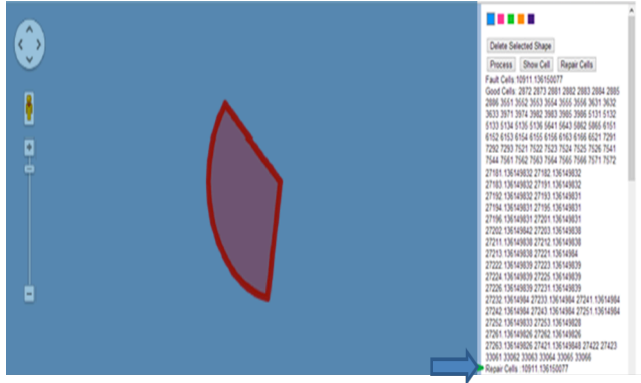


Fig. 10. Repaired cell data.

VI. CONCLUSION AND FUTURE WORK

An algorithm has been developed to analyze bad cell data automatically during the experiment on a particular geographical area. Furthermore, a database has been created to store data into the database. A software tool is developed with XML to process the data. The data has been analyzed to get more accuracy of mobile user positing. The cellular data has stored in the MySQL database. The visualization tool can visualize data on Google map. Furthermore, the data has been analyzed due to identify good cell and bad cell data. The visualize tool displays the cell shape with data value. In addition, the cell data has been analyzed based on latitude and longitude.

The good cell data has been considered when the distance between two data points of latitude and longitude are around seventy km. Furthermore bad cell data is defined when the distance between two data points are not covered of that distance. Consequently, the system identifies bad cell and good cell and finally improves the bad cell shape into good cell shape for better mobile user positioning purposes than any other existing system.

This system takes longer time to show the data on Google map in terms of large number of data set. This can be solved by advance algorithm in future.

REFERENCES

- [1] R. Mansouri, M. Saraee, and R. Amirfattahi, "Application of Data mining in predicting cell phones Subscribers Behavior Employing the Contact pattern," 2010 International Conference on Data Storage and Data Engineering (DSDE), pp. 63–68, 2010.
- [2] M. Zhou, H. Hu, and M. Zhou, "Searching XML Data by SLCA on a MapReduce Cluster," 2010 4th Intern Universal Communication Symposium (IUCS), pp. 84 – 89, 2010.
- [3] E. Neidhardt, A. Uzun, U. Bareth, and A. Kupper, "Estimating Locations and Coverage Areas of Mobile Network Cells based on Crowdsourced Data," 2013 6th Joint IFIP Wireless and Mobile Networking Conference (WMNC), pp. 1–8, 2013.
- [4] R. Baeza-Yates, C. Barrera, and V. Herskovic, "A visual tool for querying and exploring XML data," 2012 Eighth Latin American Web Congress (LA-WEB), pp. 24–27, 2012.
- [5] J. D. Mankowitz and A. J. Paverd, "Mobile Device-Based Cellular Network Coverage Analysis Using Crowd Sourcing," 2011 IEEE International Conference on Computer as a Tool - EUROCON, pp. 1–6, 2011.
- [6] X. Han, Z. Shi, W. Niu, K. Chen, and X. Yang, "Similarity-based Bayesian Learning from Semi-structured Log Files for Fault Diagnosis of Web Services," 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 589–596, 2010.
- [7] H. Lu and M. Wu, "An SUPL-based Location Platform Implementation in Mobile Network," IEEE International Conference on Communications Technology and Applications (ICCTA), pp. 284–287, 2009.
- [8] H. Xiaoyu and W. Shunxiang, "Vista Event Log File Parsing Based on XML Technology," 4th International Conference on Computer Science & Education (ICCSE), pp. 1186 – 1190, 2009.
- [9] X. Ni, "Design and Implementation of WEB Log Mining System," International Conference on Computer Engineering and Technology (ICCET), pp. 425–427, 2009.
- [10] C. W. Leong, W. Zhuang, Y. Cheng, and L. Wang, "Optimal Resource Allocation and Adaptive Call Admission Control for Voice/Data Integrated Cellular Networks," IEEE Transactions on Vehicular Technology 55(2):654-669, 2006.
- [11] E. Bertino and E. Ferrari, "XML and data integration," IEEE Internet Computing 5(6): 75–76, 2001.
- [12] C. L. C. Wong, M. C. Lee, and R. K. W. Chan, "GSM-based Mobile Positioning using WAP," 2000 IEEE Wireless Communications and Networking Conference (WCNC), pp. 874–878, 2000.
- [13] J. Yang, A. Varshavsky, H. Liu, Y. Chen, and M. Gruteser, "Accuracy Characterization of Cell Tower Localization," 12th ACM International Conference on Ubiquitous Computing (UbiComp), pp. 223–226, 2010.
- [14] M. Ulm, P. Widhalm, and N. Brandle, "Characterization of mobile phone localization errors with OpenCellID data," 4th IEEE International Conference on Advanced Logistics and Transport (ICALT), pp. 100–104, 2015.
- [15] E. Neidhardt, A. Uzun, U. Bareth, and A. Kupper, "Estimating Locations and Coverage Areas of Mobile Network Cells based on Crowdsourced Data," 2013 6th

Joint IFIP Wireless and Mobile Networking Conference (WMNC), pp. 1–8, 2013.

[16] M. Mikami, M. Miyashita, and H. Yoshino, “A Cell Identification Performance Improvement in Co-Channel

Heterogeneous Cellular Networks with Cell Range Expansion,” 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), pp. 1–5, 2015.